

Continuous-time finite-horizon ADP for automated vehicle controller design with high efficiency

Ziyu Lin

*School of Vehicle and Mobility
Tsinghua University
Beijing, China
linzy17@mails.tsinghua.edu.cn*

Jingliang Duan

*School of Vehicle and Mobility
Tsinghua University
Beijing, China
dj115@mails.tsinghua.edu.cn*

Shengbo Eben Li

*School of Vehicle and Mobility
Tsinghua University
Beijing, China
lishbo@tsinghua.edu.cn*

Haitong Ma

*School of Vehicle and Mobility
Tsinghua University
Beijing, China
maht19@mails.tsinghua.edu.cn*

Yuming Yin

*School of Vehicle and Mobility
Tsinghua University
Beijing, China
yinyuming89@gmail.com*

Bo Cheng

*School of Vehicle and Mobility
Tsinghua University
Beijing, China
chengbo@tsinghua.edu.cn*

Abstract—The design of an automated vehicle controller can be generally formulated into an optimal control problem. This paper proposes a continuous-time finite-horizon approximate dynamic programming (ADP) method, which can synthesis off-line near-optimal control policy with analytical vehicle dynamics. Lying on the general Policy Iteration framework, it employs value and policy neural networks to approximate the mappings from the system states to value function and control inputs, respectively. The proposed method can converge to the near-optimal solution of the finite-horizon Hamilton-Jacobi-Bellman (HJB) equation. We further applied our algorithm to the simulation of automated vehicle control for the path tracking maneuver. The results suggest that the proposed ADP method can obtain the near-optimal policy with 1% error and less calculation time. What is more, the proposed ADP algorithm is also suitable for nonlinear control systems, where ADP is almost 500 times faster than the nonlinear MPC ipopt solver.

Index Terms—Automated Vehicle Control, Approximate Dynamic Programming, Continuous-time Systems, Finite-horizon HJB

I. INTRODUCTION

Automated vehicles have promising potentials in safety, fuel economy, and traffic efficiency [1], [2]. As one of the critical components in the automated driving systems, motion control directly dominates a vehicle's driving performance. At present, the optimal motion controller has become an indispensable capability for high-level automated vehicles. Considering the capability in handling system nonlinearities and constraints, Model Predictive Control (MPC) is the top of the list for controller design [3]. MPC predicts the future evolution of the system and solves the control inputs online in a receding horizon fashion.

This study is supported by International Science & Technology Cooperation Program of China under 2019YFE0102200, Beijing NSF with JQ18010. Special thanks should be given to TOYOTA for their support on this study. Ziyu Lin and Jingliang Duan contribute equally to this work. All correspondences should be sent to S. Li with email: lisb04@gmail.com.

Consequently, MPC has received growing attention in the field of autonomous driving. Borrelli *et al.* (2005) proposed the nonlinear MPC control framework to perform the double lane change with active steering based on NPSOL software package. The simulation shows that under the condition of real-time computation, MPC can only obtain suboptimal solutions [4]. Verschueren *et al.* (2014) built on the generalized Gauss-Newton method to handle the real-time nonlinear MPC and got calculation time by an experiment on a Real-time Debian system [5]. Borrelli *et al.* (2014) presented the robust invariant set to satisfy robust MPC constraints of states and inputs, and did experiment on dSPACE [6]. Gerdes *et al.* (2016) proposed MPC with the constraint-based approach of envelope control, which ensures vehicle safety and brings the shared control scheme [7]. Siampis (2017) solved the nonlinear MPC problem with the Real-Time Iteration scheme and the Primal-Dual Interior-Point method, and experimented on dSPACE too [8]. Kayacan (2018) presented a real-time fast iteration scheme implemented for the nonlinear MPC framework and compared the performance and computational complexity with a linear MPC framework [9]. Batkovic *et al.* (2019) took the constraints of moving pedestrians into consideration for motion planning and control with MPC, which was also validated on Volvo XC90 [10]. This algorithm sacrifices solving precision to ensure real-time performance.

Existing MPC optimizers usually adopt the online calculation, which can not satisfy policy accuracy and the millisecond-level time requirements of onboard standard controllers, simultaneously. Taking the computing power of the Audi A8 L3 level autopilot controller, zFAS, as the benchmark, the calculation time of MPC algorithms mentioned above ([6], [10], [8], [9], [5]) is 203.4 ms, 103.6 ms, 98.9 ms, 98.4 ms, and 41.6 ms respectively. In general, the computing time assigned to the automated vehicle motion control task is less than 10ms due to the limited computing capability. The fastest single-step computing time of the five MPC algorithms listed

is 41.6 ms, which is obviously difficult to meet the controller's real-time requirements.

At present, the methods with off-line calculation have been expected to improve MPC calculation time. Approximate Dynamic Programming (ADP) method seeks approximated control policy off-line, and then apply it for online execution as a low-dimensional parameterized function [11]. The usage of ADP could significantly reduce the online computing burden. Its basic principle is to solve Hamilton-Jacobi-Bellman (HJB) equation off-line by implementing the policy iteration framework [12]–[14]. Meanwhile, existing methods transform continuous-time systems into discrete-time families, represented by MPC. Practically, the vehicle system works in a continuous-time way, which preserves significant merits in theoretical analysis and controller synthesis. In conclusion, solving the continuous-time finite horizon optimal control problem analytically with high efficiency becomes an urgent task. However, there are few existing ADP techniques on solving the optimal solution for finite-horizon control, especially for continuous-time systems. It is intractable to solve the nonlinear partial differential HJB equation.

This paper proposes a continuous-time finite-horizon ADP algorithm as the fast computation technique for automated vehicles motion control. It can handle the general nonlinear systems with known dynamic. It employs value and policy neural networks to approximate the mappings from the system states to value function and control inputs, respectively, lying on the policy iteration framework. We demonstrate our algorithm with simulations on the vehicle-tracking control problem. The main contribution can be summarized as follows:

- (1) We propose a finite-horizon variant of ADP algorithm, which can be used to find a nearly optimal policy off-line for a given continuous-time nonlinear system.
- (2) The online operation efficiency of the off-line policy is 500 times faster than the nonlinear MPC solver.

The paper is organized as follows. Section II provides the formulation of the general optimal control problem for the vehicle controller, followed by the derivative of the reshaped HJB equation. Section III introduces the general finite-time continuous-time ADP framework with parameterized function. Section IV demonstrates the effectiveness by the simulations with a linear-quadratic system and a nonlinear system. Section V concludes the paper.

II. PROBLEM FORMULATION

In this section, we introduce the linear vehicle model used in the vehicle controller design, and then the formulation for the vehicle-tracking controller is built. Subsequently, we propose the finite-horizon version of the HJB equation.

A. Vehicle model

The two degrees-of-freedom linear model is adopted to describe the vehicle dynamic shown in Fig. 1. The continuous-

time vehicle dynamic equations in an inertial frame can be derived as:

$$\begin{aligned}\dot{\vartheta} &= r, \\ \dot{r} &= \frac{a^2 k_1 + b^2 k_2}{I_{zz} v_x} r + \frac{a k_1 - b k_2}{I_{zz} v_x} v_y - \frac{a k_1}{I_{zz}} \delta, \\ \dot{v}_y &= \left(\frac{a k_1 - b k_2}{m v_x} - v_x \right) r + \frac{k_1 + k_2}{m v_x} v_y - \frac{k_1}{m} \delta.\end{aligned}\quad (1)$$

Mathematical notations are listed in Table I.

TABLE I
VEHICLE PARAMETERS NOTATION

Inertial heading angle	ϑ
Yaw rate at the center of gravity (CG)	r
Constant longitudinal velocity	v_x
Distance from mass center to front & rear axle	a, b
Front and rear wheel cornering stiffness	k_1, k_2
Mass	m
Moment of inertia around the z-axis	I_{zz}

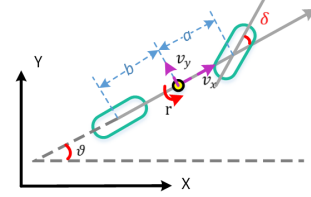


Fig. 1. The simplified vehicle dynamical model.

This classical bicycle model formulates the vehicle dynamics in reference to an initial coordinate system. However, for the vehicle tracking control problem, it is favorable to predict future vehicle motion considering the known reference trajectory. Hence, we introduce two new system states shown in Fig. 2, which are the position error d and heading angle error φ with respect to the assumed trajectory in the inertial coordinates:

$$\begin{aligned}d &= (y - y_r) \cos \vartheta_r, \\ \varphi &= \vartheta - \vartheta_r,\end{aligned}\quad (2)$$

where position error d is the distance between the center of gravity (CG) and the assumed reference trajectory. It should be noted that the assumed reference trajectory is the tangent line passing the current vehicle preview point $(x, y_r(x))$. y_r and θ_r are Y position and heading angle of the current preview point on reference trajectory, respectively. y_r is a reference trajectory function depending on vehicle position x . x, y are the coordinates of the vehicle CG in an absolute inertial frame (X, Y) . Therefore, the vehicle absolute position y is given by:

$$y = y_r + d / \cos \vartheta_r \quad (3)$$

With $\dot{\vartheta}_r = 0$, the following differential equations are satisfied:

$$\begin{aligned}\dot{d} &= v_x \sin \varphi + v_y \cos \varphi, \\ \dot{\varphi} &= \dot{\vartheta} - \dot{\vartheta}_r = r.\end{aligned}\quad (4)$$

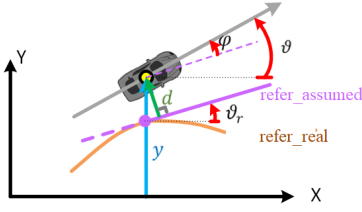


Fig. 2. Illustration of position and heading angle error.

We consider the vehicle trajectory tracking control problem with linear vehicle dynamic. The system states are $[d, \varphi, r, v_y]$ and control input is δ , listed in Table II. We assume the system is stable when x belongs to compact set Ω . With the assumption that φ is small, the bicycle model's lateral dynamic in the inertial frame is described as follows:

$$x = \begin{bmatrix} d \\ \varphi \\ r \\ v_y \end{bmatrix}, \dot{x} = Ax + Bu, B = \begin{bmatrix} 0 \\ 0 \\ -\frac{ak_1}{I_{zz}} \\ -\frac{k_1}{m} \end{bmatrix}, \quad (5)$$

$$A = \begin{bmatrix} 0 & v_x & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{a^2k_1+b^2k_2}{I_{zz}} & \frac{ak_1-bk_2}{I_{zz}v_x} \\ 0 & 0 & \frac{ak_1-bk_2}{mv_x} - v_x & \frac{k_1+k_2}{mv_x} \end{bmatrix}$$

TABLE II
STATE AND CONTROL INPUT

State	Distance between CG & trajectory	d
	The heading angle between the vehicle & trajectory	φ
	Yaw rate at the center of gravity (CG)	r
	Lateral velocity	v_y
Input	Steer angle of front wheel	δ

B. Vehicle tracking problem

Generally, the design of the automated vehicle controller can be formulated into an optimal continuous-time control problem. The basic idea of MPC is to use the vehicle dynamic model to predict the vehicle tracking system's future motion. Based on the prediction, at each time step t , the utility function is designed to minimize the tracking errors:

$$l(x(\tau), u(\tau)) = Qd^2 + R\delta^2, \quad (6)$$

where Q and R denote penalty coefficient of distance error and steer angle, respectively.

Hence, the optimal control problem (OCP) for $\forall x \in \Omega, \forall t \in [0, T]$ is given by:

$$V^*(x(t), t) = \min_u V(x, t) = \min_u \int_t^T l(x(\tau), u(\tau)) d\tau \quad (7)$$

s.t. $\dot{x} = f(x, u)$,

where T is the final time, $t \in [0, T]$ is initial time, τ is the virtual time in the future horizon, $u \in \mathbb{R}^n$ is the control input (action, in other words), $x \in \Omega \subset \mathbb{R}^m$ is the state, V is the

state value function, $l(\cdot, \cdot)$ is the utility function ($l(\cdot, \cdot) > 0$ except for the equilibrium).

C. Reshape of HJB equation

The basic principle of continuous-time finite-horizon OCP is to seek a policy $u = \pi(x, t)$ to minimize $V(x, t)$ for $\forall x$. For this purpose, we first introduce the Hamilton function:

$$H(x, u, V^\pi(x, t)) = l(x, u) + \frac{\partial V^\pi(x, t)}{\partial x^\top} f(x, u). \quad (8)$$

The Hamiltonian satisfies the following the finite-horizon self-consistency condition

$$H(x, u, V^\pi(x, t)) = -\frac{\partial V^\pi(x, t)}{\partial t}, \quad \forall t \in [0, T]. \quad (9)$$

We have deduced that $\frac{\partial V^\pi(x, t)}{\partial t} = -l(x(T), u(T))$, where $x(T)$ and $u(T)$ are obtained from $x(t)$ according to policy $\pi(x, t)$. The optimal solution satisfies the following finite-horizon HJB equation:

$$\min_u \left\{ l(x, u) + \frac{\partial V^*(x, t)}{\partial x^\top} f(x, u) \right\} = l(x^*(T), u^*(T)). \quad (10)$$

The optimal control $\pi^*(x, t)$ for $\forall x \in \Omega, \forall t \in [0, T]$ can be derived as

$$\pi^*(x, t) = \arg \min_u H \left(x, u, \frac{\partial V^*(x, t)}{\partial x^\top} \right). \quad (11)$$

It is known that the optimal control policy for the vehicle control system is the solution of finite-horizon HJB equation (10), which is the sufficient and necessary condition of optimality. There are two variables $V^*(x, t)$ and $\pi^*(x, t)$. However, HJB equation is a typical nonlinear partial differential equation, which is generally difficult to be analytically solved.

III. FINITE-HORIZON ADP ALGORITHM

In this section, we propose the continuous-time (CT) finite-horizon ADP algorithm. It is an iterative algorithm to find the nearly optimal solution of the HJB equation (10), which involves computation cycles between policy evaluation based on (9) and policy improvement (11). The value function and policy approximation can be parameterized by neural networks.

A. Basic principle

The formulation of MPC is a useful tool to understand the basic principle of ADP. MPC and ADP share very similar receding horizon control procedure, yet differs on how to find optimal action. ADP first computes optimal policy off-line and then applies it online. Differently, MPC calculates optimal action in the way of receding horizon optimization. Hence, ADP can efficiently reduce online computation since it moves most of the computation to the off-line stage.

We can always divide the implementation of OCP into two steps: (a) objective optimization, and (b) action implementation. ADP minimizes the objective function off-line and implements optimal action online. This two-step procedure can also be understood from the viewpoint of receding horizon

control, i.e., define a problem in virtual time and use optimal policy in real time, shown in Fig. 3.

For this algorithm, at each time step t , a performance index is optimized with respect to a sequence of future steering moves in order to follow the given trajectory. The first of such optimal moves is the control action applied to the plant at time t . At time $t + 1$, a new optimization is solved over a shifted prediction horizon.

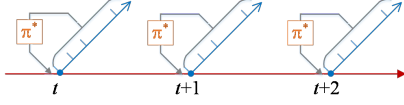


Fig. 3. Receding horizon control for ADP execution.

B. Iterative ADP Framework

The key point for the proposed algorithm is to solve HJB equation (10) iteratively. Lying on general policy iteration framework, it iteratively performs two steps shown in Fig. 4:

(1) Policy Evaluation (PEV)

PEV is to solve the value function $V^\pi(x, t)$, which is the solution of the self-consistency condition in (9). Given a policy $\pi^k(\cdot)$, the $V^\pi(x, t)$ can be solved by

$$H\left(x, u, \frac{\partial V^\pi(x, t)}{\partial x^\top}\right) = l(x(T), u(T)). \quad (12)$$

In addition, $V^\pi(0) = 0$ is an initial condition when solving (8). Here, Hamiltonian H is a little different from true Hamiltonian, in the fact that optimal value $V^*(x, t)$ is replaced with estimated value $V^\pi(x, t)$.

(2) Policy Improvement (PIM)

PIM is to find a better policy $\pi^{k+1}(x, t)$ by minimizing the "weak" HJB equation. Here, "weak" means that optimal $V^*(x, t)$ in true Hamiltonian is replaced by estimated $V^\pi(x, t)$ at each ADP iteration.

$$\pi^{k+1}(x, t) = \arg \min_u \left\{ H\left(x, u, \frac{\partial V^\pi(x, t)}{\partial x^\top}\right) \right\} \quad (13)$$

where $V^\pi(\cdot)$, x, t are known, and $\pi^{k+1}(\cdot)$ is unknown to be solved.

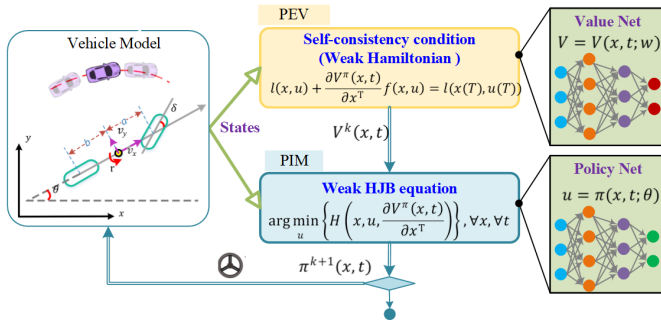


Fig. 4. CT finite-horizon ADP with parameterized function

C. Value Function and Policy Approximation

To get the nearly optimal solution quickly, both value function and policy are parameterized by the approximation functions:

$$\begin{aligned} V(x, t) &\cong V(x, t; w), \\ u &\cong \pi(x, t; \theta). \end{aligned} \quad (14)$$

The value function (i.e., critic) is a parameterized function with parameter w , and the policy (i.e., actor) is a parameterized function with parameter θ . Value function and policy are updated by designing the loss functions using the gradient descent method. The operation of each approximate function is described below.

1) Parameterized Value Function Updating rule

The critic is to minimize the following average square error based on self-consistency condition at k -step iteration PEV:

$$J_{\text{Critic}} = \mathbb{E}_{x, t \sim d_{x, t}} \left\{ \frac{1}{2} (H - l(x(T), u(T)))^2 \right\}. \quad (15)$$

where $d_{x, t}$ denotes the state-time pair distribution over $x \in \Omega$ and $t \in [0, T]$. The true-gradient of critic function equals to:

$$\frac{\partial J_{\text{Critic}}}{\partial w} = \mathbb{E}_{x, t \sim d_{x, t}} \left\{ (H - l(x(T), u(T))) \frac{\partial \left(\frac{\partial V_{w^k}}{\partial x^\top} f \right)}{\partial w} \right\}. \quad (16)$$

2) Parameterized Policy Updating rule

The parameterized actor is to minimize weak HJB equation:

$$\begin{aligned} J_{\text{Actor}} &= \mathbb{E}_{x, t \sim d_{x, t}} \left\{ H\left(x, \pi(x, t; \theta^k), \frac{\partial V(x, t; w^k)}{\partial x^\top}\right) \right\} \\ &= \mathbb{E}_{x, t \sim d_{x, t}} \left\{ l(x, \pi(x, t; \theta^k)) + \frac{\partial V_{w^k}}{\partial x^\top} f(x, \pi(x, t; \theta^k)) \right\}. \end{aligned} \quad (17)$$

The gradient of J_{Actor} is

$$\frac{\partial J_{\text{Actor}}}{\partial \theta} = \mathbb{E}_{x, t \sim d_{x, t}} \left\{ \left(\frac{\partial l(x, u)}{\partial u} + \frac{\partial \left(\frac{\partial V_{w^k}}{\partial x^\top} f \right)}{\partial u} \right) \frac{\partial \pi(x, t; \theta^k)}{\partial \theta} \right\}. \quad (18)$$

The neural networks can be updated with the given update gradients, the pseudo-code of ADP in GPI framework is shown in Algorithm 1. The loop of PEV and PIM iteratively solve the parameterized value function $V(x, t; w)$ and parameterized policy $\pi(x, t; \theta)$. The iteration of PEV and PIM will switch gradually to the HJB solution. In other words, Algorithm 1 will iteratively converge to the value function, $V(x, t; w^*) = V^*(x, t)$, and optimal control policy $\pi(x, t; \theta^*) = \pi^*(x, t)$.

IV. SIMULATION AND RESULTS

To support the continuous-time finite-horizon ADP algorithm, we apply it to the simulations with linear vehicle dynamic. It is used to solve the nearly optimal policy for the vehicle tracking system. CarSim supplies the vehicle model with steering input from the controller, shown in Fig. 5. The simulation results show that it achieves tracking the desired trajectory. Meanwhile, for a linear problem, the algorithm

Algorithm 1 CT Finite-horizon ADP

Initial with arbitrary w^0, θ^0 , learning rates α_w and α_θ
repeat

Rollout from $\forall x_t \in \Omega$ with policy π_{θ^k}

Receive and store $x(T)$

Step 1 Policy evaluation:

Calculate $\frac{\partial J_{\text{Critic}}}{\partial w}$ using (16)

Update value function using

$$w^k = w^k - \alpha_w \frac{\partial J_{\text{Critic}}}{\partial w} \quad (19)$$

Step 2 Policy improvement:

Calculate $\frac{\partial J_{\text{Actor}}}{\partial \theta}$ using (18)

Update policy using

$$\theta^k = \theta^k - \alpha_\theta \frac{\partial J_{\text{Actor}}}{\partial \theta} \quad (20)$$

until w, θ converge

works as better as the traditional MPC method. However, it can be calculated off-line, which has a promising application prospect.

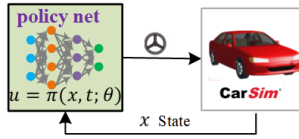


Fig. 5. Framework of Carsim verification.

A. Problem Parameters and Algorithm Details

Consider the vehicle tracking problem with linear dynamic. The desired vehicle trajectory is shown in Fig. 8. The detailed vehicle parameters are listed in Table III. The vehicle is controlled by a saturating actuator, where $\delta \in [-0.35, 0.35]$ rad. The vehicle longitudinal velocity v_x is chosen as 15 m/s.

TABLE III
VEHICLE PARAMETERS

k_1	-88000	N/rad	k_2	-94000	N/rad
a	1.14	m	b	1.4	m
m	1500	kg	I_{zz}	2420	kg · m ²

We further set the parameters in optimal control formulation, which include: the final time T is 0.5 s; the weighting matrices: $Q = 0.4$, and $R = 280$. To approximate $V(x, t; w)$ and $\pi(x, t; \theta)$, the 3-layer fully-connected NNs with 2^5 units per layer are used. For each network, the state and t input layer is followed by one hidden layer. The exponential linear units (ELUs) are chosen as activation functions. The output of value networks is $V(x, t; w)$ with softplus as the activation function. The output of policy networks is $\pi(x, t; \theta)$ with tanh function multiplied by the vector $[0.35]$, which will satisfy the

input steering angle constraint. The learning rate α_w and α_θ are both set to 0.001. We use Adam method to update the network, and the Adam update rule is used to minimize the loss functions.

B. Training Results on accuracy and efficiency

For the LQ problem, we can get the analytical solution for the discrete-time system, optimal policy, in other words.

$$U^* = [u_{0|t}^*, u_{1|t}^*, \dots, u_{T-1|t}^*]^T = -H^{-1} + Fx_t, \quad (21)$$

where,

$$H = 2(\bar{R} + \bar{S}^\top \bar{Q} \bar{S}), \quad F^\top = 2\bar{T}^\top \bar{Q} \bar{S},$$

$$\bar{S} = \begin{bmatrix} B & 0 & \dots & 0 \\ AB & B & \dots & 0 \\ \dots & \dots & \dots & \dots \\ A^{N-1}B & A^{N-2}B & \dots & B \end{bmatrix}, \quad \bar{T} = \begin{bmatrix} A \\ A^2 \\ \dots \\ A^N \end{bmatrix},$$

$$\bar{Q} = \begin{bmatrix} Q & 0 & \dots & Q \\ 0 & Q & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & P \end{bmatrix}, \quad \bar{R} = \begin{bmatrix} R & 0 & \dots & Q \\ 0 & R & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & R \end{bmatrix}.$$

The relative policy error is obtained by comparing with optimal solution $\pi^*(x, t) = u_{0|t}^*$ with the equation:

$$e_\pi = \mathbb{E}_{x, t \sim d_{x, t}} \left[\frac{\pi(x, t; \theta) - \pi^*(x, t)}{\max_{x, t \sim d_{x, t}} \pi^*(x, t) - \min_{x, t \sim d_{x, t}} \pi^*(x, t)} \right]. \quad (22)$$

The test set contains 500 states randomly selected from the compact set Ω at the beginning of each simulation. The algorithm was run ten times, and the mean of the training performance is shown in Fig. 6. We plot the policy error compared with MPC optimal solution and Hamilton of the Algorithm at each iteration. After 30000 iterations, the policy error is less than 1%, shown in Fig. 6. The result indicates that the algorithm can converge policy to optimality.

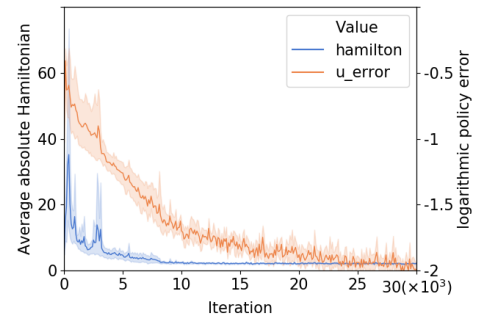


Fig. 6. Policy error and Hamilton of the Algorithm

For this optimal control problem with a linear model, we can transfer it into the linear-quadratic programming (QP) problem, and then solve it with some QP solvers, such as OSQP [15], SCS [16] and [17]. The simulations are carried on the personal computer with i5-8250U CPU. Fig. 7 compares

the calculation efficiency of ADP and the QP solvers based on the modeling language CVXPY [18] under different online applications' prediction steps. The calculation time of the QP solvers is not only longer than ADP under the same horizon but also increases faster as the number of predicted steps increases. Specially, when $N = 100$, the calculation of the proposed ADP is about 2000 times faster than the fastest QP solver OSQP. This demonstrates the effectiveness of the ADP method in linear systems control.

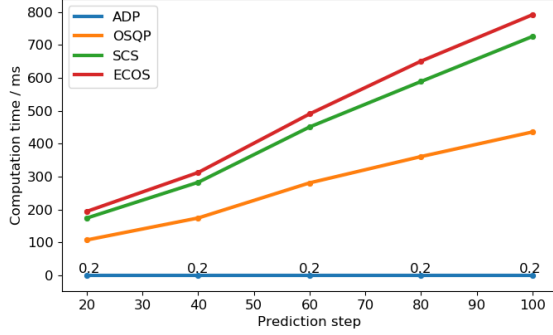


Fig. 7. Comparison of computation time of ADP and QP solvers

C. Test results on tracking problem

The vehicle trajectory controlled by the trained finite-horizon ADP algorithm is shown in Fig. 8. The learned policy can achieve following the desired trajectory quickly, which takes about 5 s in Fig. 8. It also illustrates that the algorithm can solve the continuous-time finite-horizon control problem with the linear systems.

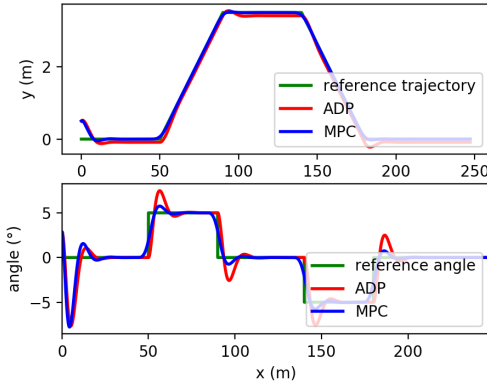


Fig. 8. Trajectory comparison with linear dynamic

Lateral position control accuracy I_{yerr} is defined as the root-mean-square value of the error between the actual position and desired position; meanwhile, $I_{y \max}$ represents maximum lateral position error. They are formulated as:

$$I_{yerr} = \sqrt{\frac{1}{N} \sum_{i=1}^N [y(i) - y_{des}(i)]^2}, \quad (23)$$

$$I_{y \max} = \max_i \{|y(i) - y_{des}(i)|\}.$$

Lateral heading angle control accuracy and maximum heading angle error can be denoted as:

$$I_{\vartheta err} = \sqrt{\frac{1}{N} \sum_{i=1}^N [\vartheta(i) - \vartheta_{des}(i)]^2}, \quad (24)$$

$$I_{\vartheta \max} = \max_i \{|\vartheta(i) - \vartheta_{des}(i)|\}.$$

Vehicle comfort can be evaluated as:

$$I_{ycomf} = \sqrt{\frac{1}{N} \sum_{i=1}^N r(i)^2}. \quad (25)$$

Here we evaluate the control performance using these indicators compared with MPC. The accuracy is similar, but the solving speed is greatly improved.

TABLE IV
COMPARISON OF CONTROL PERFORMANCE EVALUATION

	I_{yerr} (m)	$I_{y \max}$ (m)	$I_{\vartheta err}$ (rad)	$I_{\vartheta \max}$ (rad)	I_{ycomf} (rad/s)
ADP	0.013	0.52	0.0006	0.1338	0.025
MPC	0.008	0.2	0.0002	0.043	0.008

In this section, the simulation demonstrates that the proposed CT finite-horizon ADP algorithm can converge to the optimal policy for the linear system. Due to it is calculated off-line, it can improve efficiency.

D. Discussion on problems of nonlinear dynamics

We apply the CT finite-horizon ADP algorithm on a non-input-afne nonlinear vehicle system derived as in [19] [20]. The algorithm can also achieve excellent tracking performance, shown in Fig. 9. It demonstrates that the proposed ADP algorithm can also be applied to nonlinear systems effectively.

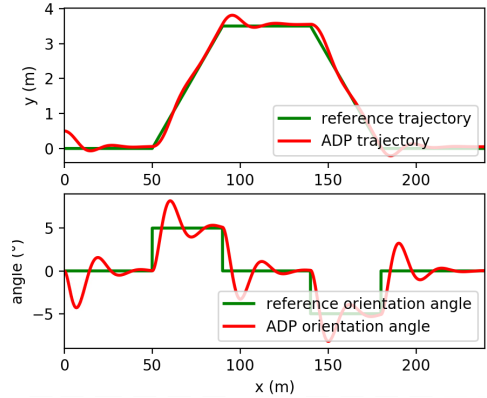


Fig. 9. Tracking performance with nonlinear dynamic

The simulation aims at control problems with finite-horizon $T = 0.5$ s. The frequency is 200 Hz. It is obvious that QP solvers listed before are not suitable for this nonlinear problem. We compare the proposed ADP algorithm with ipopt solver in CasADi framework, which is a typical open-source software for nonlinear optimal control [21]. Our simulation is carried on

personal computer with i7-8850H CPU. The results show that the average calculation time of ADP algorithm is 0.2 ms in Fig. 10, yet nonlinear MPC requires nearly 101 ms in Fig. 11. It is easy to conclude that the proposed ADP algorithm is almost 500 times faster than nonlinear MPC algorithm.

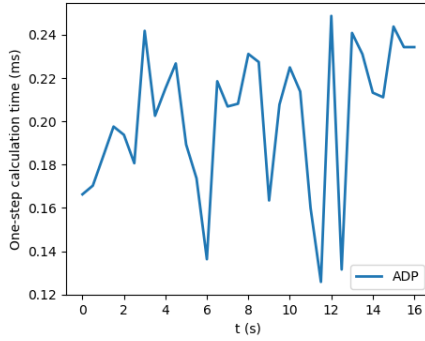


Fig. 10. Nonlinear ADP calculation time

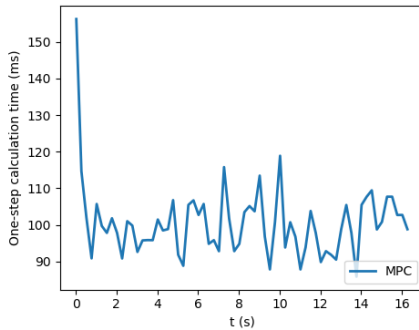


Fig. 11. Nonlinear MPC calculation time

V. CONCLUSION

The paper presents the continuous-time finite-horizon ADP algorithm for solving the general optimal control problem with known linear and nonlinear dynamics. We further verify the efficiency and optimality by the simulation on vehicle tracking problems. The result shows that it can calculate the off-line control policy for vehicle tracking problems with guaranteed optimality. For the linear control problem, the policy error is less than 1%. For the nonlinear problem, the one-step calculation time of ADP is 500 times faster than the nonlinear MPC ipopt solver.

ACKNOWLEDGMENT

We would like to acknowledge Zhengyu Liu, Hao Sun and Yarong Wang for their valuable suggestions for this research.

REFERENCES

[1] L. D. Burns, "A vision of our transport future," *Nature*, vol. 497, no. 7448, pp. 181–182, 2013.

[2] J. Duan, S. E. Li, Y. Guan, Q. Sun, and B. Cheng, "Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data," *IET Intelligent Transport Systems*, vol. 14, no. 5, pp. 297–305, 2020.

[3] S. Li, K. Li, R. Rajamani, and J. Wang, "Model predictive multi-objective vehicular adaptive cruise control," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 3, pp. 556–566, 2010.

[4] F. Borrelli, P. Falcone, T. Keviczky, J. Asgari, and D. Hrovat, "Mpc-based approach to active steering for autonomous vehicle systems," *International journal of vehicle autonomous systems*, vol. 3, no. 2–4, pp. 265–291, 2005.

[5] R. Verschuere, S. De Bruyne, M. Zanon, J. V. Frasch, and M. Diehl, "Towards time-optimal race car driving using nonlinear mpc in real-time," in *53rd IEEE conference on decision and control*. IEEE, 2014, pp. 2505–2510.

[6] Y. Gao, A. Gray, A. Carvalho, H. E. Tseng, and F. Borrelli, "Robust nonlinear predictive control for semiautonomous ground vehicles," in *2014 American Control Conference*. IEEE, 2014, pp. 4913–4918.

[7] S. M. Erlien, S. Fujita, and J. C. Gerdes, "Shared steering control using safe envelopes for obstacle avoidance and vehicle stability," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 441–451, 2015.

[8] E. Siampis, E. Velenis, S. Gariuolo, and S. Longo, "A real-time nonlinear model predictive control strategy for stabilization of an electric vehicle at the limits of handling," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 6, pp. 1982–1994, 2017.

[9] E. Kayacan, W. Saeys, H. Ramon, C. Belta, and J. M. Peschel, "Experimental validation of linear and nonlinear mpc on an articulated unmanned ground vehicle," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 5, pp. 2023–2030, 2018.

[10] I. Batkovic, M. Zanon, M. Ali, and P. Falcone, "Real-time constrained trajectory planning and vehicle control for proactive autonomous driving with road users," in *2019 18th European Control Conference (ECC)*. IEEE, 2019, pp. 256–262.

[11] P. Werbos, "Approximate dynamic programming for realtime control and neural modelling," *Handbook of intelligent control: neural, fuzzy and adaptive approaches*, pp. 493–525, 1992.

[12] P. J. Werbos, "Building and understanding adaptive systems: A statistical/numerical approach to factory automation and brain research," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 17, no. 1, pp. 7–20, 1987.

[13] J. Duan, S. E. Li, Z. Liu, M. Bujarbaruah, and B. Cheng, "Generalized policy iteration for optimal control in continuous time," *arXiv preprint arXiv:1909.05402*, 2019.

[14] J. Duan, Z. Liu, S. E. Li, Q. Sun, Z. Jia, and B. Cheng, "Deep adaptive dynamic programming for nonaffine nonlinear optimal control problem with state constraints," *arXiv preprint arXiv:1911.11397*, 2019.

[15] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, "Osqp: An operator splitting solver for quadratic programs," *Mathematical Programming Computation*, pp. 1–36, 2020.

[16] B. Odonoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *Journal of Optimization Theory and Applications*, vol. 169, no. 3, pp. 1042–1068, 2016.

[17] A. Domahidi, E. Chu, and S. Boyd, "Ecos: An socp solver for embedded systems," in *2013 European Control Conference (ECC)*. IEEE, 2013, pp. 3071–3076.

[18] S. Diamond and S. Boyd, "Cvxpy: A python-embedded modeling language for convex optimization," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2909–2913, 2016.

[19] J. Kong, M. Pfeiffer, G. Schildbach, and F. Borrelli, "Kinematic and dynamic vehicle models for autonomous driving control design," in *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2015, pp. 1094–1099.

[20] R. Li, Y. Li, S. E. Li, E. Burdet, and B. Cheng, "Driver-automation indirect shared control of highly automated vehicles with intention-aware authority transition," in *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2017, pp. 26–32.

[21] J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, "Casadi: a software framework for nonlinear optimization and optimal control," *Mathematical Programming Computation*, vol. 11, no. 1, pp. 1–36, 2019.